

Вы не должны обманывать себя: Фейнман, нейроразнообразие и честный ИИ в цифровом психическом здоровье

Источник: Frontiers in Digital Health

Оригинал: <https://www.frontiersin.org/articles/10.3389/fdgth.2026.1819411>

клиническая поддержка

нейроразнообразие

предвзятость алгоритмов

цифровое психическое здоровье

этика ИИ

Цифровое психическое здоровье претерпевает стремительную трансформацию благодаря внедрению систем искусственного интеллекта (AI — Artificial Intelligence), включая разговорных агентов, инструменты прогнозирования рисков и платформы поддержки принятия клинических решений. Несмотря на растущий энтузиазм, в данной области до сих пор не выработался единый операционный стандарт того, что представляет собой честный и заслуживающий доверия AI в психиатрическом и психологическом контекстах — стандарт, который бы явно учитывал ограничения, «слепые зоны» и группы населения, для которых внедренные системы не проходили валидацию.

В данной статье (Perspective) используются основополагающий принцип научной честности Ричарда Фейнмана, сформулированный в его выпускной речи в Калифорнийском технологическом институте (Caltech) 1974 года «Наука карго-культов» (Cargo Cult Science), и размерные метафоры Карла Сагана для разработки практической, ориентированной на раскрытие информации этической базы для использования AI в цифровом психическом здоровье.

Парадигма нейроразнообразия служит критическим стресс-тестом: растущее количество эмпирических данных указывает на то, что инструменты AI, обученные на узких поведенческих и лингвистических нормах, несут в себе существенный риск неверного представления нейроотличных пользователей, создавая потенциал для скрытого повторного кодирования стигмы, ошибочной классификации и неравенства в клинических системах.

Опираясь на теорию предиктивного кодирования (predictive processing theory) и недавние эмпирические доказательства предвзятости AI в отношении нейроотличных групп населения, данная работа предлагает пять критериев «Стандарта честности Фейнмана»: ясность охвата, раскрытие ограничений по популяциям, коммуникация неопределенности, целостность ролей и партисипаторная совместная разработка (participatory co-development).

Анализ завершается перспективной повесткой исследований для клиницистов, разработчиков и политиков, утверждая, что область будет оцениваться не только по алгоритмической эффективности, но и по степени искренности в отношении того, что эти системы знают, что они упускают и чего они не могут делать.